# A Superlinearly Convergent Minimax Algorithm for Microwave Circuit Design

JOHN W. BANDLER, FELLOW, IEEE, WITOLD KELLERMANN, STUDENT MEMBER, IEEE, AND KAJ MADSEN

*Abstract*—A new and highly efficient algorithm for nonlinear minimax optimization is presented. The algorithm, based on the work of Hald and Madsen, combines linear programming methods with quasi-Newton methods and has sure convergence properties. A critical review of the existing minimax algorithms is given, and the relation of the quasi-Newton iteration of the algorithm to the Powell method for nonlinear programming is discussed. To demonstrate the superiority of this algorithm over the existing ones, the classical three-section transmission-line transformer problem is used. A novel approach to worst-case design of microwave circuits using the present algorithm is proposed. The robustness of the algorithm is proved by using it for practical design of contiguous and noncontiguous-band multiplexers, involving up to 75 design variables.

## I. Introduction

A WIDE CLASS OF microwave circuit and system design problems can be formulated as minimax optimization problems. Most commonly, the minimax functions result from lower and/or upper specifications on a performance function of interest. In practice, we form error functions at a finite discrete set of frequencies and assume that a sufficient number of sample points have been chosen so that the discrete approximation problem adequately approximates the continuous problem. This may result in a large number of minimax functions to be minimized. Therefore, a highly efficient and fast algorithm for minimax optimization is of great importance to many microwave circuit designers and engineers. It is the purpose of this paper to present such an algorithm.

The plan of the paper is as follows. In Section II, previous work in the area of nonlinear minimax optimization is briefly reviewed. The algorithm of this paper is described in more detail in Section III, where the two methods, namely, the first-order method and the approximate second-order method, are presented and the switching conditions between the two methods are given. Our attention is focused on explaining the ideas behind the algorithm and illustrating them with microwave circuit examples. A detailed discussion on the relation of the quasi-Newton iteration of the algorithm to the Powell method for nonlinear programming is given in Appendix A. Section IV contains the comparison of the present algorithm with the existing minimax algorithms using a three-section transmission-line transformer problem.

A novel approach to worst-case tolerance design of microwave circuits taking full advantage of the present algorithm is described in Section V. Previous work in this area has been concentrated on worst-case design techniques disregarding the source of the minimax functions, i.e., the discretization of a continuous problem. Our approach, which is believed to be new to the microwave tolerance design area, integrates a search technique for maxima of the response (a technique based on cubic interpolation) with the worst-case search using linearly constrained optimization.

In Section VI, an optimization procedure for practical design of contiguous- and noncontiguous-band microwave multiplexers using the present algorithm is described and illustrated by a five-channel, 11-GHz multiplexer design.

We conclude in Section VII with an assessment of the current applicability and potential impact of the algorithm in the area of microwave circuit design.

## II. Review of Minimax Algorithms

### A. Formulation of the Problem

The mathematical formulation of the linearly constrained minimax problem is the following. Let

$$f_j(\boldsymbol{x}) = f_j(x_1, \cdots, x_n), \qquad j = 1, \cdots, m$$

be a set of $m$ nonlinear, continuously differentiable functions. The vector $\boldsymbol{x} \triangleq [x_1 \quad x_2 \cdots x_n]^T$ is the set of $n$ parameters to be optimized.

We consider the optimization problem

$$\underset{\boldsymbol{x}}{\text{minimize}}\, F(\boldsymbol{x}) \triangleq \max_j \{ f_j(\boldsymbol{x}) \}$$

subject to

$$\begin{aligned} \boldsymbol{a}_i^T \boldsymbol{x} + b_i &= 0, \qquad i = 1, \cdots, l_{\text{eq}} \\ \boldsymbol{a}_i^T \boldsymbol{x} + b_i &\geq 0, \qquad i = l_{\text{eq}} + 1, \cdots, l \end{aligned} \tag{1}$$

where $\boldsymbol{a}_i$ and $b_i$, $i = 1, \cdots, l$, are constants.

### B. Methods Based on Linearization

Over the past 15 years, this type of problem has been considered by many researchers. Usually, only the unconstrained minimax problem is treated, however. But, in the

Manuscript received February 13, 1985; revised May 28, 1985.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants A7239 and G1135.

J. W. Bandler and W. Kellermann are with the Simulation Optimization Systems Research Laboratory and the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada L8S 4L7.

K. Madsen is with the Institute for Numerical Analysis, Technical University of Denmark, Building 303, DK-2800 Lyngby, Denmark.

J. W. Bandler is also with Optimization Systems Associates, 163 Watson's Lane, Dundas, Ontario, Canada L9H 6L1.

0018-9480/85/1200-1519$01.00 ©1985 IEEE

type of methods to be described in the present paper, it is no complication and computationally costless to add the linear constraints.

Many of the minimax papers use the objective function

$$\hat{F}(x) \triangleq \max_j |f_j(x)|$$

instead of $F$. There is no significant difference between these two optimization problems. We prefer (1) since it is notationally easier and more general.

One of the earliest methods for solving the minimax problem was that of Osborne and Watson [1]. At the $k$th iterate $x_k$, their method uses a linear approximation of the nonlinear minimax problem, namely,

$$\underset{h}{\text{minimize}}\, \bar{F}(x_k, h) \triangleq \max_j \left\{ f_j(x_k) + f_j'(x_k)^T h \right\} \quad (2)$$

where $f_j'(x_k)$ denotes the gradient vector of $f_j$ w.r.t. $x$ at the point $x_k$. The minimizer $h_k$ of (2) is found using linear programming and it is used in a line search. This method may be efficient but often it is inefficient. No convergence can be guaranteed and the method can even provide convergence to a nonstationary point. Madsen [2] incorporated trust regions in the Osborne and Watson method. The linearized problem (2) is solved subject to a local bound on the variable $h$, the bound being adjusted during the iteration. No line search is used. This method has been proved to provide convergence to the set of stationary points and has a quadratic final rate of convergence when the solution is regular (Madsen and Schjaer-Jacobsen [3]). However, the rate of convergence may be very slow on singular problems.

The method of Anderson and Osborne [4] is very similar to that of Madsen. The main difference lies in the way of bounding the step length $\|h_k\|$. A different approach was used by Bandler and Charalambous [5]. They presented an approach utilizing efficient unconstrained gradient minimization techniques in conjunction with least $p$th objective functions employing extremely large values of $p$. Charalambous and Conn [6] apply an active set strategy to obtain a direction for a line search.

All of these methods are first-order methods, i.e., the search is based on first-order derivatives only. Therefore, all of these methods have problems with singular solutions and the rate of convergence may be very slow.

### C. Methods Using Second-Order Information

In order to overcome this problem, some second-order (or approximate second-order) information must be used. Hettich [7] was the first who proposed doing this. He used a Newton iteration for solving a set of equations which expresses the necessary condition for an optimum (see (6) below). However, Hettich's method is only local. It is required that the initial point is close to the solution and that the set of active functions (and constraints) is known. Han [8] suggested using nonlinear programming techniques for solving the minimax problem. He uses a nonlinear programming formulation of the minimax problem which is solved via successive quadratic programming (Powell [9]). A line search is incorporated using the minimax objective function as merit function. Overton [10] uses an approach similar to Han's but solves equality constrained quadratic problems and uses a specialized line search.

The method of Watson [11] is very similar to the method of this paper. It switches between a first- and a second-order method. The main differences between our and Watson's method are the following. Watson requires the user to provide exact first- and second-order derivatives, whereas we only require first-order derivatives. Furthermore, Watson fails to define a suitable set of criteria for switching between the first- and the second-order method. Our method has guaranteed convergence to the set of stationary points, whereas Watson's method has no such property. It can even provide convergence to a nonstationary point.

The algorithm of this paper is based on the work of Hald and Madsen [12]. It is a combination of the first-order method of Madsen [2] and an approximate second-order method. The first-order method provides fast convergence to the neighborhood of a solution. If this solution is singular, then the rate of convergence becomes very slow and a switch is made to the other method. Here a quasi-Newton method is used to solve a set of nonlinear equations that necessarily hold at a local solution of (1). This method has superlinear final convergence (see Appendix B). Several switches between the two methods may take place and the switching criteria ensure the global convergence of the combined method. Notice that the user of this algorithm is required to supply function values and first-order derivatives, whereas the necessary second derivative estimates are generated by the algorithm.

We show in Appendix A that, in the neighborhood of a local minimum of (1), our method generates the same points as the method of Powell [9] and Han [8]. However, in the latter method, a quadratic program must be solved in every iteration, whereas we have to solve only a set of linear equations or, if the solution is regular, a linear programming problem. Therefore, the computational effort used per iteration with our method is normally much smaller.

## III. Description of the Algorithm

The algorithm is a combination of two methods denoted Method 1 and Method 2. Method 1 is intended to be used far away from a solution, whereas Method 2 is a local method. We first describe these two methods.

### A. Method 1

This is essentially the algorithm of Madsen [2]. At the $k$th step, a feasible approximation $x_k$ of a solution of (1) and a local bound $\Lambda_k$ are given. In order to find a better estimate of a solution, the following linearized problem is solved:

$$\underset{h}{\text{minimize}}\, \bar{F}(x_k, h) \triangleq \max_j \left\{ f_j(x_k) + f_j'(x_k)^T h \right\}$$

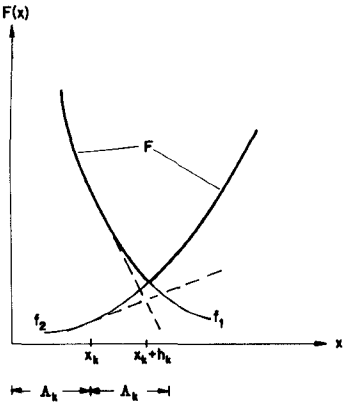



Fig. 1. An example with one variable and two functions illustrating a Method 1 iteration of the algorithm.

subject to

$$\|\boldsymbol{h}\|_\infty \leqslant \Lambda_k$$
$$\boldsymbol{a}_i^T(\boldsymbol{x}_k+\boldsymbol{h})+b_i=0, \qquad i=1,\cdots,l_{\text{eq}}$$
$$\boldsymbol{a}_i^T(\boldsymbol{x}_k+\boldsymbol{h})+b_i\geqslant 0, \qquad i=(l_{\text{eq}}+1),\cdots,l. \tag{3}$$

The solution of (3), denoted $\boldsymbol{h}_k$, is found by linear programming. Notice that $\boldsymbol{x}_k+\boldsymbol{h}_k$ is feasible. The next iterate is $\boldsymbol{x}_k+\boldsymbol{h}_k$ provided this point is better than $\boldsymbol{x}_k$ in the sense of $F$, i.e., if $F(\boldsymbol{x}_k+\boldsymbol{h}_k)<F(\boldsymbol{x}_k)$. Otherwise $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k$. In Fig. 1, an example with one variable, two functions, and no constraints ($l=0$) is shown. $F(\boldsymbol{x})$ is the kinked bold-faced curve. At $\boldsymbol{x}_k$, linear approximations of the two functions $f_1$ and $f_2$ are made and the solution of (3) is $\boldsymbol{h}_k$, which is found at the intersection of the two linear approximations. We assume that the local bound $\Lambda_k$ is so large that it has no influence. The new point is $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{h}_k$, which is seen to be close to a local minimum of $F$.

The local bound $\Lambda_k$ is introduced because the linear model (3) is a good approximation of (1) only in some neighborhood of $\boldsymbol{x}_k$. Therefore, it makes sense to consider only small values of $\|\boldsymbol{h}\|$ in connection with the linear model (3). The size of the bound is adjusted in every iteration based on a comparison between the decrease in the objective function $F$ and the decrease predicted by the model (3). If the ratio between the two is small

$$F(\boldsymbol{x}_k)-F(\boldsymbol{x}_k+\boldsymbol{h}_k)\leqslant 0.25\left[\bar{F}(\boldsymbol{x}_k,\boldsymbol{0})-\bar{F}(\boldsymbol{x}_k,\boldsymbol{h}_k)\right] \tag{4}$$

then the bound is decreased, $\Lambda_{k+1}=\Lambda_k/4$. Otherwise, if

$$F(\boldsymbol{x}_k)-F(\boldsymbol{x}_k+\boldsymbol{h}_k)\geqslant 0.75\left[\bar{F}(\boldsymbol{x}_k,\boldsymbol{0})-\bar{F}(\boldsymbol{x}_k,\boldsymbol{h}_k)\right] \tag{5}$$

then $\Lambda_{k+1}=2\Lambda_k$. If neither (4) nor (5) hold, then we leave the bound unchanged, $\Lambda_{k+1}=\Lambda_k$.

Experiments have shown that the algorithm is rather insensitive to small changes in the constants used in the updating of the bound. This method has safe global convergence properties (Madsen [2]), and if the solution is regular, then the final rate of convergence is quadratic (Madsen and Schjaer-Jacobsen [13]).

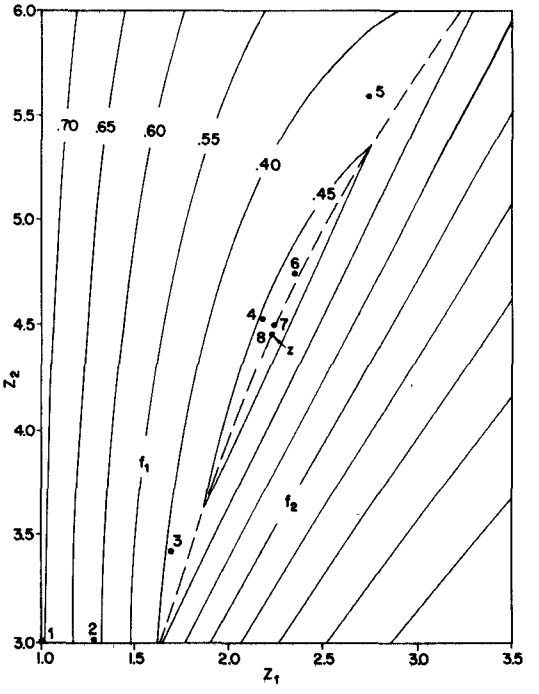



Fig. 2. Two-dimensional singular minimax problem arising from optimization of a two-section 10:1 transmission-line transformer with optimization parameters $Z_1$ and $Z_2$. The first 6 iterations are performed using Method 1 of the algorithm. Iterations 7 and 8 are performed using Method 2. The total number of iterations (function evaluations) to reach the solution with the accuracy of $10^{-6}$ is 11 (0.49 s on Cyber 170/815). If Method 2 is not used, 25 iterations (1.14 s of CPU time) are required to reach the solution.

When the solution is singular, however, the final convergence can be very slow. Consider the example of Fig. 2 in two variables where two functions are active at the solution $\boldsymbol{z}$(i.e., $f_j(\boldsymbol{z})=F(\boldsymbol{z})$ for two values of $j$). Fig. 2 shows contours for a two-section transmission-line transformer problem, where the minimax functions correspond to the reflection coefficient sampled at 11 normalized frequencies w.r.t. to 1 GHz $\{0.5,0.6,\cdots,1.4,1.5\}$. The optimization variables are characteristic impedances $Z_1$ and $Z_2$. Section lengths $l_1$ and $l_2$ are kept constant at their optimal value $l_q$, which is the quarter wavelength at the center frequency. According to Madsen and Schjaer-Jacobsen [3], this is a singular problem. Above the dotted line, $F$ is equal to one of the functions $f_j$, $F(\boldsymbol{x})=f_1(\boldsymbol{x})$, and below the dotted line, $F$ is equal to another function, $F(\boldsymbol{x})=f_2(\boldsymbol{x})$. At the dotted line, $f_1(\boldsymbol{x})=f_2(\boldsymbol{x})=F(\boldsymbol{x})$, and this line represents the bottom of a valley.

If $f_1$ and $f_2$ are different, then there is a kink at the bottom of the valley and a method based on linearization, such as Method 1, will provide fast convergence to this kink, as illustrated by the iterands in the figure (see point number 3). After the dotted line has been reached, however, the convergence towards $\boldsymbol{z}$ can be slow because the iterands have to follow a curve which passes the solution $\boldsymbol{z}$ in a smooth manner (with no kink). In the example, over eight iterations are needed to converge to a region close to the solution. Therefore, a method based on first derivatives only is not sufficient, in general, to give fast convergence. Some (approximate) second-order information is needed.

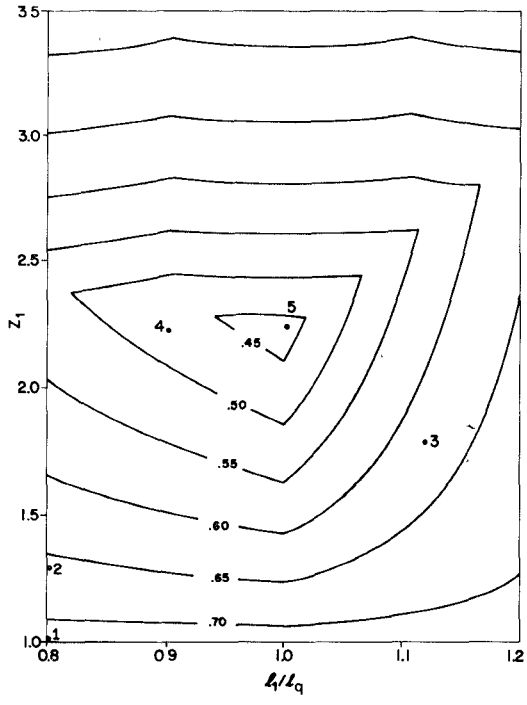



Fig. 3. Two-dimensional regular minimax problem arising from optimization of a two-section 10:1 transmission-line transformer with optimization parameters $l_1/l_q$ and $Z_1$. The first 5 iterations shown are performed using Method 1. The total number of iterations to reach the solution with the accuracy of $10^{-6}$ is 8 (0.37 s of CPU time on Cyber 170/815).

Notice that if three functions were equal at a minimum of a two-dimensional problem, then no smooth curve through the solution exists and Method 1 provides fast (quadratic) convergence to the solution.

Fig. 3 shows contours for the same two-section transformer problem. However, the optimization variables are now $l_1/l_q$ and $Z_1$. Characteristic impedance $Z_2$ and section length $l_2/l_q$ are kept at their optimal values ($l_2/l_q = 1$, $Z_2 = 4.472136$). Here, the problem is regular and five iterations are sufficient to reach the vicinity of the solution.

### B. Method 2

It is a local method. It is assumed that a point near a solution $z$ is known, and that the active sets $A(z) \triangleq \{j | f_j(z) = F(z)\}$ and $C(z) \triangleq \{i | a_i^T z + b_i = 0\}$ are known. At a local minimum $z$ of (1), the following necessary conditions hold (see, e.g., [7]):

$$\begin{aligned} \sum_{j \in A(z)} \lambda_j f_j'(z) - \sum_{i \in C(z)} \mu_i a_i &= \mathbf{0} \\ \sum_{j \in A(z)} \lambda_j - 1 &= 0 \\ f_{j_0}(z) - f_j(z) &= 0, \qquad j \in A(z) \backslash \{j_0\} \\ a_i^T z + b_i &= 0, \qquad i \in C(z), \end{aligned} \tag{6}$$

where the multipliers $\lambda_j$ and $\mu_i$ are nonnegative and $j_0 \in A(z)$ is fixed. Method 2 is an approximate Newton method for solving the nonlinear system (6) (in the variables $(z, \lambda, \mu)$). Exact first derivatives are used but the matrix $\Sigma \lambda_i f_j''(z)$ is approximated using a modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) update (see Appendix A for details). In this way, an approximate Jacobian $J_k$ is obtained at the estimate $(x_k, \lambda^{(k)}, \mu^{(k)})$ of the solution of (6). The next estimate is found by

$$J_k \begin{bmatrix} \Delta x_k \\ \Delta \lambda^{(k)} \\ \Delta \mu^{(k)} \end{bmatrix} = -R\left(x_k, \lambda^{(k)}, \mu^{(k)}\right)$$

$$\left(x_{k+1}, \lambda^{(k+1)}\right), \mu^{(k+1)}\right) = \left(x_k, \lambda^{(k)}, \mu^{(k)}\right) + \left(\Delta x_k, \Delta \lambda^{(k)}, \Delta \mu^{(k)}\right) \tag{7}$$

where $R(z, \lambda, \mu) = \mathbf{0}$ is the vector formulation of (6).

### C. The Combined Method

The combined method is the algorithm which we recommend and use in this paper. Initially, Method 1 is used and the active sets used in (6) are estimated. When a singular local minimum seems to be approached, a switch to Method 2 is made. If the Method 2 iteration is unsuccessful, Method 1 is used again. Several switches between the two methods may take place. When Method 1 is used, we say that the iteration is in Stage 1, otherwise it is in Stage 2. A detailed description of the two stages follows.

*The Stage 1 Iteration:* We have a point $x_k$, a local bound $\Lambda_k$, and a matrix $J_k$ which should approximate the Jacobian of (6).

1) $x_{k+1}$ and $\Lambda_{k+1}$ are found using Method 1, and approximations $A_{k+1}$ and $C_{k+1}$ of the active sets at $x_{k+1}$ are found via the active sets at the solution $h_k$ of the linear model problem (3).

2) An estimate $(\lambda^{(k+1)}, \mu^{(k+1)})$ of the multipliers is found through a least-squares solution of (6) with $(x_{k+1}, A_{k+1}, C_{k+1})$ inserted for $(z, A(z), C(z))$. This estimate is used for finding a new Jacobian estimate $J_{k+1}$ by the BFGS update as described in Appendix A.

3) A switch to Stage 2 is made if the following two conditions hold:

a) The active set estimates have been constant over three consecutive different Stage 1 iterates.
b) The components of $\lambda^{(k+1)}$ and $\mu^{(k+1)}$ are nonnegative.

*The Stage 2 Iteration:* $x_k, \Lambda_k, J_k$, and active set estimates $A_k, C_k$ are given.

1) Find $(x_{k+1}, \lambda^{(k+1)}, \mu^{(k+1)})$ and $J_{k+1}$ using Method 2 with $(A_k, C_k)$ inserted for $(A(z), C(z))$.

2) Let $A_{k+1} = A_k$, $C_{k+1} = C_k$, and $\Lambda_{k+1} = \Lambda_k$.

3) Switch to Stage 1 if one of the following conditions holds:

a) A function or constraint outside of $A_{k+1}$ or $C_{k+1}$ is active at $x_{k+1}$.
b) A component of $\lambda^{(k+1)}$ or $\mu^{(k+1)}$ is negative.
c) $\|R(x_{k+1}, \lambda^{(k+1)}, \mu^{(k+1)})\| > 0.999 \|R(x_k, \lambda^{(k)}, \mu^{(k)}\|$ (see (7) for the definition of $R$).

TABLE I
COMPARISON OF ALGORITHMS FOR THE THREE-SECTION TRANSFORMER (NUMBER OF FUNCTION EVALUATIONS)

| Algorithm | Starting Point $\mathbf{x}_0^1$ | Starting Point $\mathbf{x}_0^2$ |
|---|---|---|
| This algorithm[1] | 18 * | 21 ** |
| Hald and Madsen [12] | 21 | 46 |
| Agnew [19] | Alg I 23 | 64 |
| | Alg II 20 | 109 |
| Bandler and Charalambous [20] | 95 | 155 |
| Charalambous and Conn [21] | 162 | 67 |
| Conn[22] | 67 | 80 |
| Madsen [2] | 253 | 707 |
| Madsen and Schjaer-Jacobsen [3] | 29 | 48 |
| This algorithm[2] | 15 + | 22 ++ |

Execution times on Cyber 170/815 in seconds are * 0 6, ** 0 7, + 0 57, ++ 0 85

"Active" frequency points selected by the cubic interpolation search
0 50000, 0 76999, 1 23001, 1 50000

[1] without cubic interpolation
[2] with cubic interpolation

This completes the description of the combined method.

It has been shown [12] that the combined method can only converge to stationary points and that the final rate of convergence is quadratic on regular problems and superlinear on singular problems (provided that the Jacobian of (6) is regular).

The results published by Hald and Madsen [12] correspond to the combined method as described here except that the PSB (Powell's symmetric Broyden) formula was used for updating $J_k$ in Method 2. Our numerical results indicate that the use of the BFGS formula as described in Appendix A is significantly better (see Table I).

For this paper, we have used the MMLC version [14] of the algorithm based on the earlier implementation due to Hald [15].

## IV. COMPARISON WITH OTHER ALGORITHMS

### A. The Test Problem

To compare the performance of the present algorithm to other minimax algorithms, a three-section 100-percent relative bandwidth 10:1 transmission-line transformer problem has been chosen (see Fig. 4). It is a special case of an $N$-section transmission-line transformer. Originally studied by Bandler and Macdonald [16], [17], this type of test problem is now widely considered.

The problem is to minimize the maximum reflection coefficient of this matching network. A detailed discussion on the formulation of direct minimax response objectives is presented in [18].

Formally, the problem is to

$$\underset{x}{\text{minimize}}\, F(x) = \max_{[0.5,1.5]} |\rho(x,\omega)| \tag{8}$$

where

$$x = \left[l_1/l_q \quad Z_1 \quad l_2/l_q \quad Z_2 \quad l_3/l_q \quad Z_3\right]^T.$$

The minimax functions represent the modulus of the reflection coefficient sampled at the 11 normalized frequencies $\omega$ (w.r.t. 1 GHz) {0.5, 0.6, 0.7, 0.77, 0.9, 1.0, 1.1, 1.23, 1.3, 1.4, 1.5}. The known quarter-wave solution is given by $l_1 = l_2 = l_3 = l_q$, $Z_1 = 1.63471$, $Z_2 = 3.16228$, $Z_3 = 6.11729$, where $l_q$ is the quarter wavelength at the center frequency, namely,

$$l_q = 7.49481 \text{ cm for 1 GHz.}$$

The corresponding maximum reflection coefficient is 0.19729. Two starting points have been used

$$x_0^1 = [0.8 \quad 1.5 \quad 1.2 \quad 3.0 \quad 0.8 \quad 6.0]^T$$

$$x_0^2 = [1.0 \quad 1.0 \quad 1.0 \quad 3.16228 \quad 1.0 \quad 10.0]^T.$$

Gradient vectors with respect to section lengths and characteristic impedances are obtained using the adjoint network method.

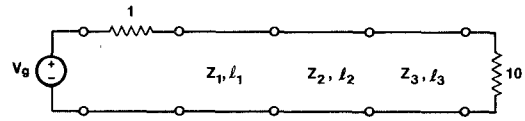



Fig. 4. Three-section, 10:1 transmission-line transformer used as a test problem to compare the performance of minimax algorithms.

### B. Performance of the New Algorithm

Table I shows the performance of the new algorithm as compared to other algorithms. Table I also shows results obtained using the present algorithm with a cubic interpolation search for maxima of the response. Using this technique, the number of sample points can be reduced from 11 to 4, and we do not have to know in advance the location of frequency points corresponding to the maxima of the response. More information on the cubic interpolation search technique is given in Section V in the context of a new approach to worst-case design of microwave circuits.

To show the influence of the parameters *DX* (initial step length of the iterative algorithm) and *KEQS* (the number of successive iterations with identical sets of active residual functions that is required before a switch to Stage 2 is made), the optimization has been performed several times for different values of *DX* and *KEQS*. The resulting numbers of residual function evaluations required to achieve the accuracy $EPS = 10^{-6}$, as well as the number shifts to Stage 2 are summarized in Table II (the numbers of shifts are given in parentheses).

It can be observed that the increasing values of *KEQS* correspond to slightly increased numbers of function evaluations. Moreover, too small and too large values of *DX* require more residual function evaluations because of adjustments which are performed by the algorithm. From other experiments, it was observed that the increasing values of *KEQS* correspond, generally, to smaller numbers of shifts to Stage 2 (some too early shifts are eliminated).

## V. WORST-CASE NETWORK DESIGN

### A. Preliminary Remarks

In this section, we will formulate the fixed tolerance problem (FTP) [23], [24] embodying a worst-case search and the selection of sample points for the discrete ap-

TABLE II
THE INFLUENCE OF THE CONTROLLING PARAMETERS DX AND KEQS ON THE NUMBER OF FUNCTION EVALUATIONS

| | KEQS | | |
|---|---|---|---|
| DX | 2 | 3 | 4 |
| 0 1 | 21(2) | 23(2) | 24(2) |
| 0 25 | 19(2) | 18(2) | 19(2) |
| 0 5 | 18(2) | 20(2) | 22(2) |
| 0 75 | 18(2) | 18(2) | 20(2) |
| 1 0 | 21(2) | 22(2) | 23(2) |

proximation of a continuous problem. As mentioned in the introduction, the discretization of a continuous problem may result in a large number of minimax functions to be minimized. The size of the problem increases dramatically if we want to consider tolerances on network parameters since, for each frequency point selected to represent the response $2^n$ ($n$ is the number of network parameters), minimax functions have to be created if we want to consider all vertices of the tolerance region.

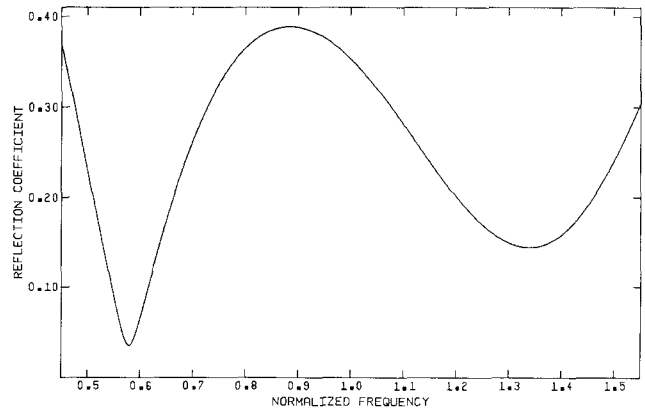
A number of methods have been proposed for solving the worst-case problem. Schjaer-Jacobsen and Madsen [24] suggest the application of interval arithmetic. Bandler *et al.* [25] and Tromp [26] described methods which rely on the assumption that the functions considered are one-dimensionally convex.

Our approach to the fixed tolerance problem is a double iterative algorithm. For each outer iteration of minimization, first a search using cubic interpolation is done to determine frequency points which are candidates for active functions, and then a number (equal to the number of selected minimax functions) of inner loop optimizations are performed to determine the worst case for each of the minimax functions.

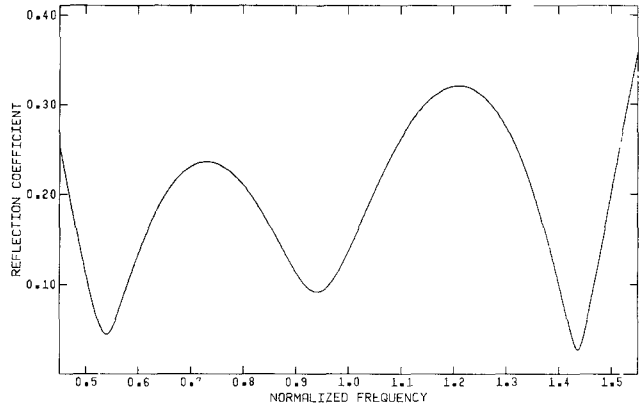
The advantage of our approach is that the worst-case search (done by means of linearly constrained optimization) and the actual minimization are linked together such that each worst-case calculation affects immediately the outer iteration of minimization.

### *B. Cubic Interpolation Search Technique*

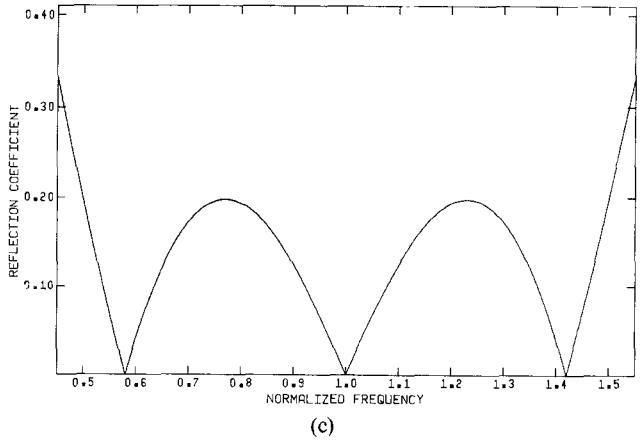
The cubic interpolation technique allows us to consider the minimum number of frequency points to adequately approximate the continuous problem. In many cases, the discretization of a continuous problem may not be adequate to give the continuous minimax solution. As illustrated in Fig. 5, the solution obtained using uniformly spaced sample points may not be optimal in the continuous minimax sense since some of the peaks of the response (or error function) have been missed. One way to overcome this difficulty is to use densely spaced sample points. This, however, may result in a prohibitively large number of minimax functions to be optimized. Therefore, it is desirable to develop a technique to locate the maxima of the response w.r.t. frequency and to track these maxima during the optimization process as they shift along the frequency




(a)




(b)




(c)

Fig. 5. (a) Response of the three-section transmission-line transformer at the starting point $x_1^0$. The initial sample points are 0.5, 0.8, 1.2, and 1.5. The uniformly spaced (with the step length of 0.1) search points are 0.5, 0.6,..., 1.5. The sample points selected by the algorithm using cubic interpolation are 0.5, 0 884, 1.2, and 1.5. The edges of the frequency interval are kept as the sample points. The initial sample point 0.8 has been replaced by the point 0.884 since a maximum has been detected between 0.8 and 0.9. The sample point 1.2 has not been changed since no maximum has been found between points 1.2 and 1.3. (b) Response of the three-section transmission-line transformer after the first iteration of optimization. The sample points selected by the cubic interpolation search (with the step length of 0.1) are 0.5, 0.729, 1.210, and 1.5. Now the initial sample point 1.2 has been replaced by 1.210 since a maximum has been detected between points 1.2 and 1.3. The sample point 0.884 has been replaced by a new sample point 0.729 resulting from the maximum between 0.7 and 0.8. (c) Response of the three-section transformer at the solution. The sample points selected by the cubic interpolation search (with the step length of 0.1) before the last iteration of optimization are 0.5, 0.770, 1.230, and 1.5.

axis due to the changes in the values of optimization parameters. Such a technique has been developed by Bandler and Chen [27]. It is based on the cubic interpolation formulas of Fletcher and Powell [28]. For convenient reference, the formulas are given in Appendix C.

### C. Fixed Tolerance Problem

To present the method for worst-case tolerance design, we will introduce some notation which basically follows that of [23].

A design consists of design data of the nominal point $\phi^0 \triangleq [\phi_1^0 \quad \phi_2^0 \cdots \phi_n^0]^T$ and a set of associated tolerances $\epsilon \triangleq [\epsilon_1 \quad \epsilon_2 \cdots \epsilon_n]^T$, where $n$ is the number of network parameters. An outcome of a circuit is any point $\phi \triangleq [\phi_1 \; \phi_2 \cdots \phi_n]^T$ such that $\phi_i = \phi_i^0 + \epsilon_i \mu_i$, $-1 \leqslant \mu_i \leqslant 1$, $i = 1,2,\cdots,n$. The tolerance region $R_\epsilon$ is defined as $R_\epsilon \triangleq \{\phi | \phi_i = \phi_i^0 + \epsilon_i \mu_i,\ i = 1,2,\cdots,n\}$. The extreme points of $R_\epsilon$ are called the vertices and are obtained by setting $\mu_i = \pm 1$.

We consider a set of $m$ nonlinear functions

$$f_j(\phi^0) \triangleq f(\phi^0, \omega_j), \qquad j \in J \triangleq \{1,2,\cdots,m\} \tag{9}$$

where $\omega_j$, $j \in J$, is an independent parameter (frequency). The number of functions $m$ is equal to

$$m = m_{\max} + 2$$

where $m_{\max}$ is the number of the maxima of the response and 2 represents the edges of the frequency interval $[\omega_l, \omega_u]$.

The fixed tolerance problem can be defined on the basis of the worst-case objective function [24] as that of determining

$$\min_{\phi^0} F(\phi^0) = \min_{\phi^0} \max_{j \in J} \left\{ \max_{\phi \in R_\epsilon} f_j(\phi) \right\}. \tag{10}$$

For each outer iteration of minimization w.r.t. $\phi^0$, $m$ frequency points are determined (by a search technique based on cubic interpolation) and $m$ linearly constrained optimizations are performed to find the worst cases.

At the $k$th outer iteration of minimization, we have an approximation $\phi_k^0$ of the solution and we solve $m$ linearly constrained optimizations, where the $j$th problem, $j \in J$ is

$$\underset{\phi_k}{\text{minimize}} \left( -f_j(\phi_k) \right) \tag{11}$$

subject to

$$\left(\phi_i^0\right)_k - \epsilon_i \leqslant (\phi_i)_k \leqslant \left(\phi_i^0\right)_k + \epsilon_i, \qquad i = 1,2,\cdots,n.$$

Once $\phi_k^*$ for the $j$th function is determined, we can identify whether the worst-case occurred at a vertex using the following criteria.

Let

$$(y_i)_k = \left| \left(\phi_i^0\right)_k - \left(\phi_i^*\right)_k \right|. \tag{12}$$

If $|(y_i)_k - \epsilon_i| \leqslant 10^{-5}$, then the worst-case occurred at a vertex, for which $\mu_i$, $i = 1,2,\cdots,n$, are easy to determine

$$\mu_i = \begin{cases} -1, & \text{if} \quad \left(\phi_i^*\right)_k \leqslant \left(\phi_i^0\right)_k \\ +1, & \text{otherwise} \end{cases}. \tag{13}$$

TABLE III
FIXED TOLERANCE PROBLEM FOR THE THREE-SECTION 10:1 TRANSFORMER

| | | |
|---|---|---|
| Number of Minimax Functions | 4 | |
| Number of Variables | 6 | |
| Required Accuracy of the Solution | $10^{-5}$ | |
| Assumed Tolerances | 5% | |
| Step Size in the Cubic Interpolation Search | 0.1 | |
| Solution Vector | $\ell_1/\ell_q = 0.96373$ | $Z_2 = 3.22493$ |
| | $Z_1 = 1.67797$ | $\ell_3/\ell_q = 0.96483$ |
| | $\ell_2/\ell_q = 0.98720$ | $Z_3 = 6.04817$ |
| "Active" Frequency Points | 0.50000, 0.78726, 1.27242, 1.50000 | |
| Maximum Refl. Coefficient | 0.33589 | |
| Number of Function Evaluations | 32 | |
| Execution Time on Cyber 170/815 (in seconds) | 8.1 | |

The function values $f_j$, $j \in J$, and the gradients of $f_j$, $j \in J$, which are returned to the outer iteration are evaluated at a point $(\phi_j^*)_k$, i.e., were the $j$th worst case occurred.

### D. Illustration of the Approach

The three-section transmission-line transformer is used to illustrate the approach and its validity for worst-case design. Numerical results are summarized in Table III. As expected, the nominal parameter values are different from the values obtained for the nominal design problem. The location of the two internal maxima of the response has also changed as compared to the nominal design problem. Each linearly constrained optimization to determine the worst-case for the particular frequency with the accuracy $10^{-3}$ requires about four iterations of the algorithm.

## VI. CONTIGUOUS AND NONCONTIGUOUS-BAND MULTIPLEXER DESIGN

### A. Introductory Remarks

Practical design and manufacture of contiguous- and noncontiguous-band multiplexers consisting of multicavity filters distributed along a waveguide manifold has been a problem of significant interest [29]–[31]. Recently, a general multiplexer design procedure using an extension of the normal least-squares method has been described [32].

We present here a general multiplexer optimization procedure exploiting exact network sensitivities. The simulation and sensitivity analysis aspects of the problem, together with a number of examples of multiplexer optimization (including a twelve-channel, 12-GHz multiplexer without dummy channels), have been described in [33]. All design parameters of interest, e.g., waveguide spacings, input–output, and filter coupling parameters, can be directly optimized. A typical structure under consideration is shown in Fig. 6.

### B. Formulation of the Problem

A wide range of possible multiplexer optimization problems can be formulated and solved by appropriately defining specifications on common port return loss and individual channel insertion loss functions. The minimax error functions are created using those specifications, simulated exact multiplexer responses, and weighting factors.

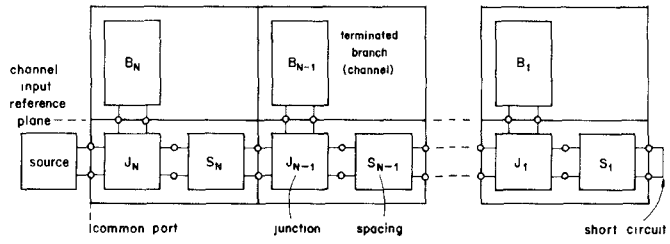



Fig. 6 The multiplexer configuration under consideration. $J_1, J_2, \ldots, J_N$ are arbitrarily defined 3-port junctions, $B_1, B_2, \ldots, B_N$ are terminated branches or channels which may each be represented in reduced cascade form, and $S_1, S_2, \cdots, S_N$ are usually waveguide spacing elements.

The minimax functions $f_j(\boldsymbol{x})$, $j \in J$, are of the form

$$w_{Uk}^1(\omega_i)\left(F_k^1(\boldsymbol{x},\omega_i) - S_{Uk}^1(\omega_i)\right) \quad (14)$$

$$-w_{Lk}^1(\omega_i)\left(F_k^1(\boldsymbol{x},\omega_i) - S_{Lk}^1(\omega_i)\right) \quad (15)$$

$$w_U^2(\omega_i)\left(F^2(\boldsymbol{x},\omega_i) - S_U^2(\omega_i)\right) \quad (16)$$

$$-w_L^2(\omega_i)\left(F^2(\boldsymbol{x},\omega_i) - S_L^2(\omega_i)\right) \quad (17)$$

where $F_k^1(\boldsymbol{x},\omega_i)$ is the insertion loss for the $k$th channel at the $i$th frequency, $F^2(\boldsymbol{x},\omega_i)$ is the return loss at the common port at the $i$th frequency, $S_{Uk}^1(\omega_i)(S_{Lk}^1(\omega_i))$ is the upper (lower) specification on insertion loss of the $k$th channel at the $i$th frequency, $S_U^2(\omega_i)(S_L^2(\omega_i))$ is the upper (lower) specification on return loss at the $i$th frequency, and $w_{Uk}^1, w_{Lk}^1, w_U^2, w_L^2$ are the arbitrary user-chosen nonnegative weighting factors.

### C. Five-Channel 11-GHz Multiplexer Design

The procedure is illustrated by designing an 11-GHz, five-channel multiplexer having the center frequencies and bandwidths (similar to those of [32]) given in Table IV.

Suppose we want to design this multiplexer such that certain specifications on the common port return loss and individual channel insertion loss functions are satisfied. A lower specification of 20 dB on return loss over the passbands of all five channels should be satisfied. We want also to control return loss between channels 1 and 2, 2 and 3, and 4 and 5 in a similar way. We impose also additional specifications on insertion loss for all channels, i.e., we want the insertion loss in the transition bands not to drop below 20 dB.

We start the design process with five identical six-pole, pseudo-elliptic function filters. Starting values of the coupling coefficients for the filters are given in the following matrix [31]:

$$M = \begin{bmatrix} 0 & 0.62575 & 0 & 0 & 0 & 0 \\ 0.62575 & 0 & 0.57615 & 0 & 0 & 0 \\ 0 & 0.57615 & 0 & 0.32348 & 0 & -0.74957 \\ 0 & 0 & 0.32348 & 0 & 1.04102 & 0 \\ 0 & 0 & 0 & 1.04102 & 0 & 1.04239 \\ 0 & 0 & -0.74957 & 0 & 1.04239 & 0 \end{bmatrix}.$$

The initial spacing lengths are set equal to $\lambda_{gk}/2$ (half the wavelength corresponding to the $k$th center frequency).

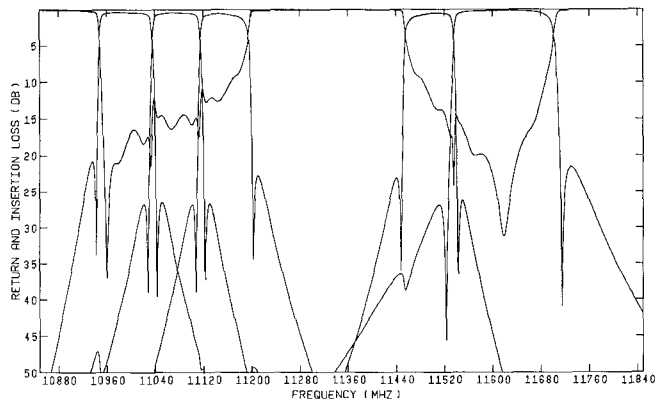



Fig. 7. Responses of the 5-channel, 11-GHz multiplexer at the starting point of the optimization process.

TABLE IV
MULTIPLEXER CENTER FREQUENCIES AND BANDWIDTHS

| Channel | Center Frequency (MHz) | Bandwidth (MHz) |
|---|---|---|
| 1 | 10992 5 | 81 |
| 2 | 11075 0 | 76 |
| 3 | 11155 0 | 76 |
| 4 | 11495 0 | 76 |
| 5 | 11618 5 | 154 |

The filters are assumed lossy and dispersive. Waveguide junctions are assumed nonideal.

Fig. 7 shows the responses of the multiplexer at the start of the optimization process. As we see, the specifications on the common-port return loss are seriously violated.

The optimization process is performed in several steps. First, we select only nonzero couplings, input/output transformer ratios, and filter spacings as optimization variables. This gives a total of 45 optimization variables. The error functions resulting from the multiplexer responses and specifications are created at 51 nonuniformly spaced frequency points. An improved design is obtained after 30 function evaluations (230 s on the Cyber 170/815). The responses corresponding to the first step of the optimization process are shown in Fig. 8.

In order to completely satisfy the design specifications we perform a second step of optimization in which we release additional optimization variables, i.e., cavity resonances. This gives a total of 75 nonlinear optimization variables. Using the same frequency points as in step 1 and results of the first optimization as a starting point, we continue the optimization process. After 30 additional

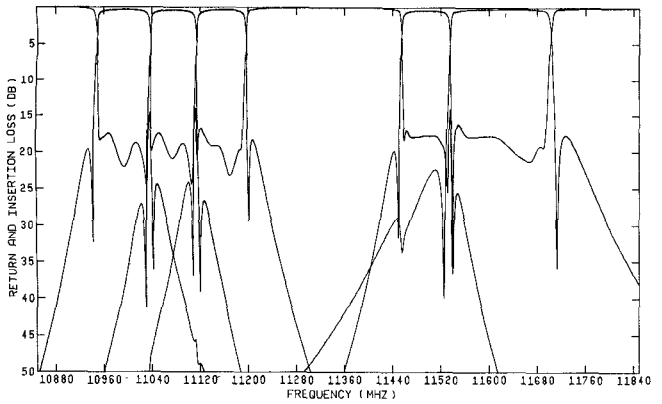



Fig. 8. Responses of the 5-channel, 11-GHz multiplexer after the first 30 iterations using 45 optimization variables and 51 nonuniformly spaced sample points.

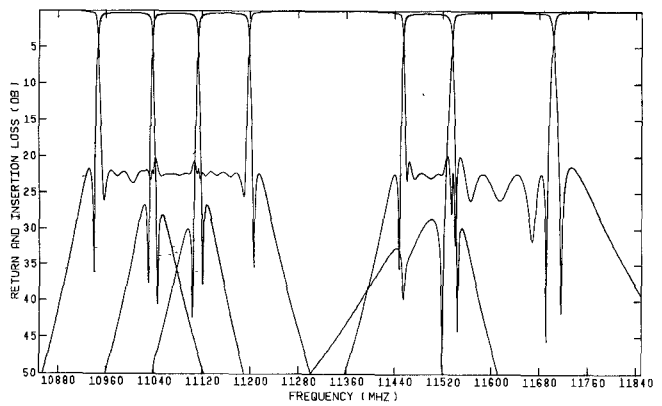



Fig. 9. Responses of the 5-channel, 11-GHz multiplexer after 30 additional iterations using 75 optimization variables and 51 minimax functions.

function evaluations (and 470 s of CPU time on the Cyber 170/815), the design specifications are satisfied and the optimized responses of the five-channel multiplexer are shown in Fig. 9. To improve the return loss response of the multiplexer, the third step of optimization is performed in which a search technique for maxima of the response is employed. This gives 66 minimax functions and the same number of variables as previously. After 25 additional function evaluations (and 360 s of CPU time on the Cyber 170/815), we obtain the final optimized responses as shown in Fig. 10.

In the approach presented, the emphasis is on achieving a maximally effective set of early iterations of optimization using a subset of all possible optimization variables. This subset should correspond to "dominant" variables of the problem. Initial selection of the variables can be facilitated by the full knowledge and experience of the designer and by an initial sensitivity analysis at selected frequency points.

## VII. Conclusions

We have described a new and highly efficient algorithm for nonlinear minimax optimization problems which arise in microwave circuit design. The algorithm combines linear programming methods with quasi-Newton methods and the convergence is at least superlinear. Comparison made

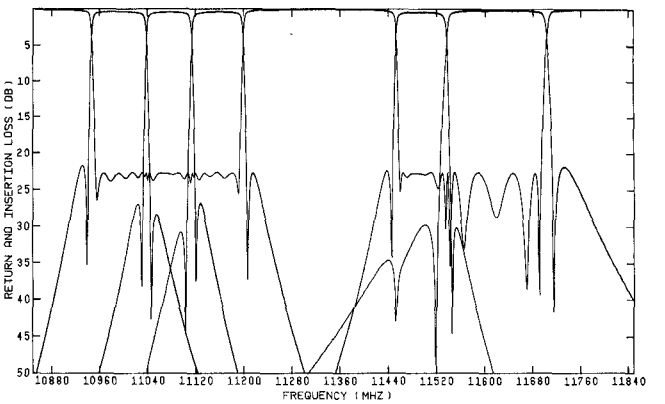



Fig. 10. The final optimized responses of the 5-channel multiplexer obtained with 75 variables and 66 minimax functions. A search technique for the peaks of the return loss response has been employed.

with the existing minimax algorithms on the classical three section transmission-line transformer problem shows clearly that this algorithm is better in terms of the number of function evaluations required to reach the solution with a desired accuracy.

We have presented a novel approach to worst-case tolerance design of microwave circuits integrating a cubic interpolation based search technique for maxima of the response with the worst-case search using linearly constrained optimization. The validity of the approach has been demonstrated by solving a fixed tolerance problem for a three-section transmission-line transformer. We emphasize that our approach does not require the designer to know in advance the location of frequency points corresponding to the maxima of the response and significantly reduces the number of sample points adequately approximating the continuous response. This aspect of our approach is particularly important since it can significantly reduce the number of minimax functions for which the worst cases have to be found.

The robustness of the algorithm presented makes possible the practical design of contiguous- and noncontiguous-band microwave multiplexers. To our knowledge, our work is the first successful attempt to use gradient-based optimization for multiplexer design, as well as being the largest nonlinear optimization process ever demonstrated on microwave circuit design for a reasonable computational cost.

We feel that the algorithm presented will have a significant impact on microwave circuit design techniques and practices allowing the designer to consider problems of greater size than usually done in the past, including tolerances on circuit parameters.

## Appendix A
## Quasi-Newton Iteration and the Powell Method

Here the details of the approximate Newton iteration used in Method 2 are given. Furthermore, it is shown that close to a local minimum, Method 2 generates the same points as Powell's sequential quadratic programming method [9] applied to the nonlinear programming formula-

tion of (1). Therefore, our method has the same local convergence properties as that of Powell.

We consider one iteration of Method 2. For simplicity, we use the notation $x = x_k$, $\lambda = \lambda_k$, $\mu = \mu_k$, $A = A(z)$, and $C = C(z)$. In a Newton iteration for solving (6), we should use the Jacobian

$$R'(x,\lambda,\mu) = \begin{bmatrix} \sum_{j\in A} \lambda_j f_j''(x) & E & -F \\ 0\,0\cdots 0 & 1\,1\cdots 1 & 0\,0\cdots 0 \\ G^T & 0 & 0 \\ F^T & 0 & 0 \end{bmatrix} \tag{A1}$$

where $E$ has the columns $f_j'(x)$, $j \in A$, $F$ has the columns $a_i$, $i \in C$, and $G$ has the columns $f_{j_0}'(x) - f_j'(x)$, $j \in A\backslash\{j_0\}$. Only the upper left-hand block involves more than first derivatives. In Method 2, this block is approximated by an updating formula, whereas the exact values are used in the other blocks of $R'$.

The Lagrangian function corresponding to (1) is

$$L(x,\lambda,\mu) = \sum_{j=1}^{m} \lambda_j f_j(x) - \sum_{i=1}^{l} \mu_i [a_i^T x + b_i] \tag{A2}$$

so the upper left-hand block of (A1) is $L_{xx}''(x,\lambda,\mu)$ since $\lambda_j = 0$ for $j \notin A$.

This block is approximated by the BFGS formula with the modifications of Powell [9] that keep the approximation positive definite. Thus, the matrix $J_k$ of (7) is

$$J_k = \begin{bmatrix} B_k & E & -F \\ 0\,0\cdots 0 & 1\,1\cdots 1 & 0\cdots 0 \\ G^T & 0 & 0 \\ F^T & 0 & 0 \end{bmatrix} \tag{A3}$$

where $B_k$ is updated through

$$B_{k+1} = B_k - B_k s s^T B_k / [s^T B_k s] + y y^T / [s^T y]$$

with

$$s = x_{k+1} - x$$
$$y = L_x'(x_{k+1},\lambda,\mu) - L_x'(x,\lambda,\mu). \tag{A4}$$

An iteration of Method 2 is now given by (A3), (7), and (A4) with

$$R(x,\lambda,\mu) = \begin{bmatrix} L_x'(x,\lambda,\mu) \\ \sum \lambda_j - 1 \\ e \\ f \end{bmatrix} \tag{A5}$$

where $e$ has the components $f_{j_0}(x) - f_j(x)$, $j \in A\backslash\{j_0\}$ and $f$ has the components $a_i^T x + b_i$, $i \in C$.

Now consider the nonlinear programming formulation of problem (1)

$$\underset{(x,\delta)}{\text{minimize}}\, G(x,\delta) \triangleq \delta$$

subject to

$$\begin{aligned} \delta - f_j(x) &\geq 0, && j = 1,\cdots,m \\ a_i^T x + b_i &= 0, && i = 1,\cdots,l_{\text{eq}} \\ a_i^T x + b_i &\geq 0, && i = (l_{\text{eq}}+1),\cdots,l. \end{aligned} \tag{A6}$$

Assume that Powell's sequential quadratic programming method is used to solve (A6). At the iterate $(x,\delta) = (x_k,\delta_k)$, the following subproblem is solved:

$$\underset{(h,p)}{\text{minimize}}\, Q(h,p) = p + h^T \bar{B}_k h$$

subject to

$$\begin{aligned} \delta + p - [f_j(x) + f_j'(x)^T h] &\geq 0, && j = 1,\cdots,m \\ a_i^T(x+h) + b_i &= 0, && i = 1,\cdots,l_{\text{eq}} \\ a_i^T(x+h) + b_i &\geq 0, && i = (l_{\text{eq}}+1),\cdots,l \end{aligned} \tag{A7}$$

with $\bar{B}_k$ being a positive definite estimate of the Hessian of the Lagrangian. $\bar{B}_k$ has the dimensions $n \times n$. Actually, it should be $(n+1)(n+1)$, but the row and column corresponding to $p$ are left out for notational convenience since they have no influence. The next iterate is $(x_{k+1},\delta_{k+1}) = (x,\delta) + \alpha(h,p)$, where $(h,p)$ denote the solution of (A7), and close to a solution $\alpha = 1$ is necessary for fast convergence. We assume that $(x,\delta)$ is close to a solution $(z,\delta^*)$ and that $\alpha = 1$. Assume further that $x$ is so close to $z$ that the active constraints at the solution of (A7) are the same as at the solution of (A6). These are identified by the indices $j \in A = A(z)$ and $i \in C = C(z)$.

Using these assumptions, we can find the solution of (A7) using the Kuhn–Tucker conditions. They give

$$\begin{bmatrix} \bar{B}_k h \\ 1 \end{bmatrix} - \sum_{j\in A} \bar{\lambda}_j \begin{bmatrix} -f_j'(x) \\ 1 \end{bmatrix} - \sum_{i\in C} \bar{\mu}_i \begin{bmatrix} a_i \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{aligned} p - [f_j(x) + f_j'(x)^T h] &= 0, && j \in A, \\ a_i^T(x+h) + b_i &= 0, && i \in C. \end{aligned} \tag{A8}$$

This is equivalent to the system

$$\begin{aligned} \bar{B}_k h + \sum_{j\in A} \bar{\lambda}_j f_j'(x) - \sum_{i\in C} \bar{\mu}_i a_i &= 0 \\ \sum_{j\in A} \bar{\lambda}_j - 1 &= 0 \\ [f_{j_0}(x) + f_{j_0}'(x)^T h] - [f_j(x) + f_j'(x)^T h] &= 0, \quad j \in A\backslash\{j_0\} \\ a_i^T(x+h) + b_i &= 0, \quad i \in C. \end{aligned} \tag{A9}$$

Using (A3), (A5), and (A2) and a small calculation, it is seen that (A9) is the same as (7) with $h = \Delta x_k$, $\bar{\lambda} = \lambda^{(k+1)}$ and $\bar{\mu} = \mu^{(k+1)}$ (provided that $\bar{B}_k = B_k$). Thus, the point $x + h$ found by Powell's method is the same as the point $x_{k+1}$ found by Method 2. Furthermore, Powell uses $\bar{\lambda}$ and $\bar{\mu}$ as the new multiplier estimates so also here there is coincidence with Method 2.

Finally, the matrices $\bar{B}_k$ and $B_k$ are updated through the same formula. This is seen from the fact that the Lagrangian of (A6) is

$$\bar{L}(w,\lambda,\mu) = \delta - \sum_{j=1}^{m} \lambda_j [\delta - f_j(x)] - \sum_{i=1}^{l} \mu_i [a_i^T x + b_i] \tag{A10}$$

with $w = (x, \delta)$. In Powell's method, $\bar{B}_k$ is updated by (A4) with $\bar{L}$ instead of $L$. Therefore, it is seen from (A2) and (A10) that the updates of $\bar{B}_k$ and $B_k$ are identical.

Consequently, Method 2 is identical to Powell's method in its final stages provided that the matrices to be updated are initialized in the same way.

## Appendix B
## Superlinear Convergence

If the sequence $\{r_k\}$ converges to $r^*$ in such a way that

$$\lim_{k \to \infty} \frac{|r_{k+1} - r^*|}{|r_k - r^*|} = \beta < 1$$

the sequence is said to converge linearly to $r^*$ with convergence ratio $\beta$.

The case where $\beta = 0$ is referred to as superlinear convergence.

## Appendix C
## Cubic Interpolation Formula

As a well-known fact, a maximum of a continuous differentiable function $e(\omega)$ is characterized by $e' \triangleq \partial e/\partial\omega = 0$ and $\partial^2 e/\partial\omega^2 < 0$. This implies a change in the sign of $\partial e/\partial\omega$ and, in the neighborhood of the maximum, $\partial e/\partial\omega$ decreases as frequency increases. It follows that if there exist two points $\omega_1 < \omega_2$ such that

$$e'_{\omega_1} > 0 \quad \text{and} \quad e'_{\omega_2} < 0$$

at least one maximum of $e(\omega)$ lies between $\omega_1$ and $\omega_2$. If $\omega_1$ and $\omega_2$ are close enough to exclude the existence of multiple maxima, the location of the detected maximum can be estimated by the cubic interpolation formula [28]

$$\omega_{\max} = \omega_2 - \frac{(\omega_2 - \omega_1)[x - y - e'_{\omega_2}]}{e'_{\omega_1} - e'_{\omega_2} + 2x} \tag{C1}$$

where

$$y = -e'_{\omega_1} - e'_{\omega_2} + 3\frac{e(\omega_2) - e(\omega_1)}{\omega_2 - \omega_1} \tag{C2}$$

and

$$x = \left[y^2 - e'_{\omega_1} e'_{\omega_2}\right]^{1/2}. \tag{C3}$$

## Acknowledgment

The authors are grateful to S. H. Chen and S. Daijavad of McMaster University for their high-level contribution to the success of this research program. They also acknowledge the contribution of Dr. W. M. Zuberek, currently with the Memorial University of Newfoundland, in the early stage of the work on worst-case circuit design. The assistance of M. L. Renault of McMaster University in preparing the numerical examples is also gratefully acknowledged.

## References

[1] M. R. Osborne and G. A. Watson, "An algorithm for minimax optimization in the nonlinear case," *Comput. J.*, vol. 12, pp. 63–68, 1969.

[2] K. Madsen, "An algorithm for minimax solution of overdetermined systems of nonlinear equations," *J. IMA*, vol. 16, pp. 321–328, 1975.

[3] K. Madsen and H. Schjaer-Jacobsen, "Singularities in minimax optimization of networks," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 456–460, 1976.

[4] D. H. Anderson and M. R. Osborne, "Discrete, nonlinear approximations in polyhedral norms: A Levenberg-like algorithm," *Num. Math.*, vol. 28, pp. 157–170, 1977.

[5] J. W. Bandler and C. Charalambous, "Practical least $p$th optimization of networks," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-20, pp. 834–840, 1972.

[6] C. Charalambous and A. R. Conn, "An efficient method to solve the minimax problem directly," *SIAM J Numerical Analysis*, vol. 15, pp. 162–187, 1978.

[7] R. Hettich, "A Newton-method for nonlinear Chebyshev approximation," in *Approximation Theory* (Lecture Notes in Mathematics, vol. 556), R. Schaback and K. Scherer, Eds. Berlin: Springer, 1976, pp. 222–236.

[8] S. P. Han, "Variable metric methods for minimizing a class of non-differentiable functions," *Mathematical Programming*, vol. 20, pp. 1–13, 1981.

[9] M. J. D. Powell, "The convergence of variable metric method for nonlinearly constrained optimization calculations," in *Nonlinear Programming*, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, Eds. New York: Academic Press, 1978, pp. 27–63.

[10] M. L. Overton, "Projected Lagrangian algorithms for nonlinear minimax and $l_1$ optimization," Stanford Univ., Dep. Comput. Sci., Rep. STAN-CS-79-752, 1979.

[11] G. A. Watson, "The minimax solution of an overdetermined system of non-linear equations." *J. IMA*, vol. 23, pp. 167–180, 1979.

[12] J. Hald and K. Madsen, "Combined LP and quasi-Newton methods for minimax optimization," *Mathematical Programming*, vol. 20, pp. 49–62, 1981.

[13] K. Madsen and H. Schjaer-Jacobsen, "Linearly constrained minimax optimization," *Mathematical Programming*, vol. 14, pp. 208–223, 1978.

[14] J. W. Bandler and W. M. Zuberek, "MMLC—A Fortran package for linearly constrained minimax optimization," Dept. Elec. and Comput. Eng., McMaster University, Hamilton, Canada, Rep. SOS-82-5, 1982.

[15] J. Hald, "MMLA1Q, a Fortran subroutine for linearly constrained minimax optimization," Institute for Numerical Analysis, Tech. Univ. Denmark, Rep. NI-81-01, 1981.

[16] J. W. Bandler and P. A. Macdonald, "Optimization of microwave networks by razor search," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-17, pp. 552–562, 1969.

[17] J. W. Bandler and P. A. Macdonald, "Cascaded noncommensurate transmission-line networks as optimization problems," *IEEE Trans. Circuit Theory*, vol. CT-16, pp. 391–394, 1969.

[18] J. W. Bandler, "Optimization methods for computer-aided design," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-17, pp. 533–552, 1969.

[19] D. Agnew, "Improved minimax optimization for circuit design," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 791–803, 1981.

[20] J. W. Bandler and C. Charalambous, "New algorithms for network optimization," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-21, pp. 815–818, 1973.

[21] C. Charalambous and A. R. Conn, "Optimization of microwave networks," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 834–838, 1975.

[22] A. R. Conn, "An efficient second order method to solve the (constrained) minimax problem," Dept. Combinatorics and Optimization, Univ. of Waterloo, Waterloo, Ontario, Rep. CORR 79-5, 1979.

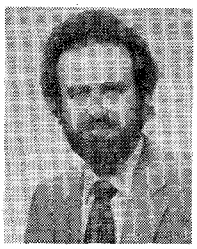
[23] J. W. Bandler, P. C. Liu, and H. Tromp, "A nonlinear programming approach to optimal design centering, tolerancing and tuning," *IEEE Trans. Circuits Syst.*, vol. CAS-23, pp. 155–165, 1976.

[24] H. Schjaer-Jacobsen and K. Madsen, "Algorithms for worst-case tolerance optimization," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 775–783, 1979.

[25] J. W. Bandler, P. C. Liu, and J. H. K. Chen, "Worst case network tolerance optimization," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 630–641, 1975.

[26] H. Tromp, "The generalized tolerance problem and worst case search," in *Proc. IEEE Conf. Computer-Aided Design of Electronic and Microwave Circuits and Systems* (Hull, England), 1977, pp. 72–77.

[27] J. W. Bandler and S. H. Chen, "Interactive optimization of multi-coupled cavity microwave filters," Dept. Elec. and Comput. Eng., McMaster University, Hamilton, Canada, Rep. SOS-84-13-R, 1984.
[28] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Computer J.*, vol. 6, pp. 163–168, 1963.
[29] A. E. Atia, "Computer-aided design of waveguide multiplexers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-22, pp. 332–336, 1974.
[30] M. H. Chen, F. Assal, and C. Mahle, "A contiguous band multiplexer," *COMSAT Tech. Rev.*, vol. 6, pp. 285–306, 1976.
[31] M. H. Chen, "A 12-channel contiguous band multiplexer at *Ku*-band," in *1983 IEEE Int. Microwave Symp. Dig.* (Boston), pp. 77–79.
[32] R. G. Egri, A. E. Williams, and A. E. Atia, "A contiguous-band multiplexer design," in *1983 IEEE Int. Microwave Symp. Dig.* (Boston), pp. 86–88.
[33] J. W. Bandler, S. H. Chen, S. Daijavad, and W. Kellermann, "Optimal design of multicavity filters and contiguous-band multiplexers," *Proc. 14th Eur. Microwave Conf.* (Liege, Belgium), 1984, pp. 863–868.



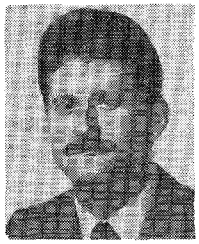
**John W. Bandler** (S'66–M'66–SM'74–F'78) was born in Jerusalem, Palestine, on November 9, 1941. He studied at Imperial College of Science and Technology, London, England, from 1960–1966. He received the B.Sc. (Eng.), Ph.D., and D.Sc. (Eng.) degrees from the University of London, London, England, in 1963, 1967, and 1976, respectively.

He joined Mullard Research Laboratories, Redhill, Surrey, England, in 1966. From 1967 to 1969, he was a Postdoctorate Fellow and Sessional Lecturer at the University of Manitoba, Winnipeg, Canada. He joined McMaster University, Hamilton, Canada, in 1969, where he is currently a Professor of Electrical and Computer Engineering. He has served as Chairman of the Department of Electrical Engineering and Dean of the Faculty of Engineering. He currently directs research, which has received substantial support by the Natural Sciences and Engineering Research Council of Canada under its Operating and Strategic Grants Awards, in the Simulation Optimization Systems Research Laboratory.

Dr. Bandler is also currently President of Optimization Systems Associates. He has provided consulting services and software to numerous organizations in the electronic, microwave, and electrical power industry, specializing in advanced applications of simulation, sensitivity analysis, and mathematical optimization techniques.

He is a contributor to *Modern Filter Theory and Design*, Wiley-Interscience, 1973. He has over 200 publications, four of which appear in *Computer-Aided Filter Design*, IEEE Press, 1973, one in *Microwave Integrated Circuits*, Artech House, 1975, and one in *Low-Noise Microwave Transistors and Amplifiers*, IEEE Press, 1981. Dr. Bandler was an Associate Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES (1969–1974). He was Guest Editor of the Special Issue of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES on Computer-Oriented Microwave Practices (March 1974).

Dr. Bandler is a Fellow of the Institution of Electrical Engineers (Great Britain) and a member of the Association of Professional Engineers of the Province of Ontario (Canada).

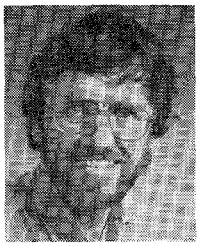


**Witold Kellermann** (S'82) was born in Opole, Poland, on August 8, 1954. He received the M.Sc. degree in electrical engineering from the Technical University of Wroclaw, Wroclaw, Poland, in 1978.

From 1978 to 1980, he was a Research and Teaching Assistant in the Department of Electrical Engineering, Higher School of Engineering, Opole, Poland. Since 1981, he has been employed as a Teaching Assistant in the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada.

At McMaster University, he joined the Simulation Optimization Systems Research Laboratory, where he has been working under the supervision of Dr. J. W. Bandler. Mr. Kellermann was awarded an Ontario Graduate Scholarship for 1984/85. He is currently completing his Ph.D. requirements in electrical engineering from McMaster University.

His research interests are in the area of optimization-based computer-aided design of circuits and systems.



**Kaj Madsen** was born in Denmark in 1943. He received his Cand. Scient. degree in mathematics from the University of Aarhus in 1968.

He has been a Lecturer in numerical analysis and computer science since 1968, apart from the academic year 1973 to 1974, when he was at AERE Harwell, Didcot, England. Most of the time he has been with the Institute for Numerical Analysis at the Technical University of Denmark, but during the years 1981–1983, he was at the Computer Science Department at Copenhagen University. During the summer of 1978, he visited McMaster University, Hamilton, Canada. His fields of interest in teaching and research are programming languages, optimization, and interval analysis.